The National Environmental Information Exchange Network

Network Nodes: A Primer

December 14, 2001 Draft

Contents

I. Background Information on the National Environmental Exchange Network

Information is fundamental to the work of environmental protection. State environmental agencies and the U.S. Environmental Protection Agency (EPA) depend on the rational flow of quality information for every aspect of their work, as individual agencies and collectively. Yet, many of the current systems and approaches to information exchange are ineffective and burdensome. Once fully established, the National Environmental Information Exchange Network (Network) will provide an alternative to the current approach. The Network will utilize Internet technologies and approaches to transform the exchange of environmental information. The specific technologies used for the Network are detailed in the *Blueprint for a National Environmental Information Exchange Network* (*Blueprint*) document which can be accessed at http://www.sso.org/ecos/eie/COMPLETE_BLUEPRINT_JUNE_01_FINAL.pdf.

The business case for the Network, as outlined in the *Blueprint*, rests on the Network's ability to provide the following value to its participants:

- The infrastructure and set of tools to efficiently establish information exchanges to meet business needs
- A framework, based on explicitly-recognized data stewardship, for informing and improving the investments agencies are making, and will continue to make, in information. And thereby…
- … Improvement in the quality and timeliness of information available to and from state agencies, EPA, and other information trading partners. Specifically, the Network will streamline – and ultimately reduce – the costs of flowing data to and from EPA, state agencies, and other stakeholders.

Comparing the status quo to the Network further illustrates the improvements offered by Network participation.

Most of the current processes for reporting data from state environmental agencies to EPA were developed decades ago. They consist of sending paper, double or triple data entry into separate state/federal systems, or using translator programs to convert information from a newer system into a legacy system's format (or vice versa). These current processes represent a substantial management cost to state agencies and EPA.

Participation in the Network hinges upon the use of Network nodes – simple web services that enable the exchange of information in an agreed-upon format using a known address on the Internet that can issue and respond to authorized requests for specific information using standard protocols.

Network data exchanges will replace (and be superior to) the traditional approach to information exchange that relied upon paper submittals, duplicate data entry, and/or various electronic submissions. Partners who choose to utilize the Network will do so in place of their traditional approaches. The core of the Network, however, is not technology; it is a commitment to change the way data is exchanged.

The Network will depend on the ability of environmental partners to negotiate and then define the exact format in which data will be exchanged, to document the exchange in Trading Partner Agreement (TPAs),

and to hold partners responsible for fulfilling these agreements. Responsibility for data quality, timeliness, format, and availability will be explicitly defined, documented, and agreed upon by a designated individual for each partner. Data originators will fulfill these agreements by maintaining information sources (nodes) on the Network that can provide information upon authorized request.

Network Steering Board and Implementation Plan

In the fall of 2001, the Interim Network Steering Group (INSG) and the State/EPA Information Management Work Group (IMWG) have proposed the establishment of a Network Steering Board (NSB) that will support the Network once the INSG sunsets in the spring of 2002.

As envisioned, the mission of the Network Steering Board (NSB) will be to not only take over the INSG's responsibilities for Network implementation and success, but also manage the administration and support of the Network to ensure that it is effective, impartial, and responsive to all Network participants.

The functions of the Network Steering Board will be

1. Oversee and steer (guide) implementation of the Network, including tracking and reporting to the IMWG on milestones and issues outlined in the Network Implementation Plan.
2. Manage the resources made available to the NSB for Network management.
3. Coordinate with IMWG Action Teams those activities that relate to Network implementation.
4. Sponsor teams as needed to develop policies and best practice recommendations on:

    a. Operation and maintenance of the Network Registry/Repository for Data Exchange Template (DET) development and management;
    b. DET creation;
    c. Use of Environmental Data Standards Council (EDSC) adopted data standards in DETs;
    d. Support of broader harmonization and consistency within DETs and with other standards;
    e. Security operations, policies, and other technical elements; and
    f. Trading Partner Agreement (TPA) guidance development, including TPA templates and checklists.

II. Frequently Asked Questions About Network Nodes

1. What is a Network node?

A Network node (node) is the hardware and software used to exchange environmental information on the Network. A node uses the Internet, a set of standard protocols, and appropriate security measures to issue and respond to authorized requests for specific information.

Technically, a Network node is a simple environmental information web service that processes authorized queries and sends the requested information in a standard format, XML (eXtensible Markup Language). A node validates this information against a predefined Schema/Data Exchange Template (DET). (These transactions are governed by TPAs). Node management approaches will probably "look" different for each agency, but will include a formal process for approving (i.e., signing TPAs for) those node flows that require them.

Nodes have the following features:

- Each Network participant has only one node, although that node may handle many kinds and types of information requests
- A node is the only route for Network delivery of official information[1]
- The node is the single place for each participant to present its standard node catalog of available information and associated Network metadata (e.g., its TPAs and description of the information). Data and associated information must be provided by a node to be on the Network
- The node is the single place where each member implements the minimal but essential transport, security, and query protocols described in the *Blueprint* and specified in a TPA, and
- The node is the only place where a member's compliance with a TPA can be demonstrated or evaluated.

2. How can Network nodes be used?

Once established, Network nodes can be used as general purpose "data exchange portals" for many agency efforts. Eventually, state environmental agencies and EPA will flow information through their nodes ranging from the following:

- Official State/EPA flows (delegated regulatory program information)
- Other State/EPA flows (such as facility data to EPA's Facility Registry System)

[1] Participants may choose to make some information on their nodes publicly available and/or to use their nodes as "back-ends" for public access websites

- Other flows from facilities to state environmental agencies and (if applicable) facilities to EPA, such as the eDMR (electronic Discharge Monitoring Report)

- State and EPA information exchange with other parties (e.g., academics, scientists, community groups, the general public, etc.)

- Information exchange among state agencies, such as water quality data for basin management planning

- Simultaneously sending public data in XML format and displaying the data on a public website in a "user-friendly format

- Currently, there are no limitations to the types of information that Network participants could exchange using their nodes.

While nodes can be used by Network participants in any way they see fit, it is expected that node information exchanges for official Network and related State/EPA flows will be undertaken according to the necessary official, agreed-upon TPAs and DETs (Schema or Document Type Definitions (DTDs)).

3. What are the advantages of creating a Network node (and in so doing, participate in the Network)? Why build a node if our current process of reporting data to EPA works?

Most current data exchanges are either manual or semi-manual operations, customized for each application. The Network will evolve this approach to a web services-based architecture that will provide a common exchange infrastructure for data exchanges. Creating Network nodes is the first step in Network participation. Nodes provide stable tools for receiving information provided by other nodes and can support automated data exchanges and information can only be on the Network if it has flowed from a node.

Specifically, node technologies and tools allow for information to be shared securely and reliably with less effort and fewer resources than was previously the case. By using one standard, vendor-neutral, publicly-available format (XML) for all exchanged data, the need to painstakingly develop new software application "translators" for each transaction will become a relic of the past. Further, nodes will allow information to be retrieved on a "come and get it" (active data retrieval) basis. (In EPA's case, this would include aggregation of the information in existing or new systems.) For example, EPA or other Network participants will be able to automate regular active data retrieval at night, and thereby more effectively use bandwidth that is currently heavily relied on during the day. This will even be true for transactions that require the strictest level of security. Therefore, when the participants wish to do so, data will be made available on a node 24 hours a day, seven days a week, negating the need to schedule transactions ahead of time. In general, establishing and using nodes, coupled with implementing the policy components of the Network, will make sharing information easier.

Finally, and most importantly, shifting away from previous methods of providing and/or sharing information will allow Network participants to focus more on information quality and using the data for environmental analysis and decision-making.

4. What is XML and why is it important?

As outlined in the Blueprint, Network nodes take advantage of the benefits provided by eXtensible Markup Language (XML) and its companion technologies. XML is a meta-markup language – a standard

format for electronic documents. Data is included in XML documents as strings of text, and the data is surrounded by text markup that describes the data. XML is called "extensible" because it can be extended and adapted to meet many different needs as defined by the user. XML is not a programming language. (Source: O'Reilly 2001).

Any platform can create or process XML-formatted information, alleviating the need to design, develop, and maintain custom translators for each information transaction. Since XML is simple in design and simple in use, XML is becoming the standard format for Internet-based information exchange. It allows information to be exchanged easily and quickly with less effort and fewer resources.

Further, most recent enterprise information management software packages, including the products from the software industry giants, are employing (incorporating) XML. The latest versions of most enterprise software include built-in XML capabilities that can be used for node functions. In addition, many software companies have developed separate XML management tools that enable users to format and send their data in XML with very little effort.

Due to these efforts, and increasing use of XML in all sectors, the costs of sharing information across the Internet are dropping. (Please see Appendix B for more information about how XML is being used.)

5. Where can I learn more about XML and related technologies?

High-level overviews of XML are available at most larger bookstores. (See also, Appendix D: Additional Resources). More detailed, technical, and up-to-date information on XML technologies, tools, guidelines, and related topic can be found at the following websites:

- **The World Wide Web Consortium (W3C) (http://www.w3.org/):** W3C was created in 1994 to lead the World Wide Web to its full potential as a forum for information, commerce, communication, and collective understanding. The Consortium develops protocols that promote the Web's evolution and ensure its interoperability. W3C has more than 500 member organizations from around the world and has earned international recognition for its contributions to the growth of the Web.
- **The Organization for the Advancement of Structured Information Standards (OASIS) (http://www.oasis-open.org/):** OASIS is a non-profit, international consortium that creates interoperable industry specifications based on public standards such as XML and SGML (need to spell out), as well as others that are related to structured information processing.
- **ebXML (http://www.ebxml.org/):** Sponsored by OASIS and UN/CEFACT (the United Nations Centre for Trade Facilitation and Electronic Business – www.unece.org/cefact), ebXML is a set of specifications that enables enterprises of any size and in any geographical location to conduct business over the Internet. Using ebXML, companies have a standard method to exchange business messages, conduct trading relationships, communicate data in common terms, and define and register business processes. ebXML has three types of products: technical reports, reference materials, and white papers.
- **XML.gov (http://www.xml.gov/):** XML.gov was created to facilitate the efficient and effective use of XML through cooperative efforts among government agencies, including partnerships with commercial and industrial organizations. The XML Working Group is charged with developing an online information resource (at xml.gov) that defines and documents an evolving strategy and a set of tasks for effective and well-coordinated usage of XML to support governmental functions.

6. Are Network nodes being built now? If so, by whom?

Several state agencies and EPA (Central Data Exchange (CDX)) are building Network nodes as part of the Network Node Pilot Project. In the spring and summer of 2001, four state agencies (in Delaware, Nebraska, New Hampshire, and Utah) participated in the Network Node Pilot Project – Alpha Phase, which resulted in the first "proof of concept" pilot nodes that demonstrated simple data transactions between nodes. The Network Node Pilot Project – Beta Phase is being implemented in the fall and winter of 2001. As part of the Beta Phase, the same states (above) plus Florida, New Mexico, and EPA (CDX), are building the next generation pilot nodes, or "Beta nodes." The Beta Phase will help bring the Network concept closer to reality and the initial development of Version 1.0 Network nodes in 2002. (See Part III for more information on the Network Node Pilot Project.)

7. How does current node work relate to future plans for Network implementation?

The INSG is developing an implementation plan which will be turned over to the Network Steering Board in early 2002. (See also, Background: page 2.) The Network Node Pilot Project – Beta Phase will also be completed in early 2002, and as such, this project's efforts will help to establish the groundwork for overall Network implementation.

8. What are TPAs and how do they relate to the development and usage of nodes?

Trading Partner Agreements (TPAs) are written agreements that define the partners, information, stewardship, security, and other items essential for the exchange of information between two or more trading partners on the Network. In short, TPAs establish formal processes for managing the flow of information across the Network. TPAs may apply to exchanges initiated by the sender or those initiated at the request of the receiver. If exchanges are intended to meet mandatory reporting requirement, TPAs are necessary when automated exchanges are to take place without operator intervention. TPAs are also advisable between any parties (e.g., state agencies and regions) that wish to establish an ongoing business process involving automated electronic exchanges of information.

TPA guidance is expected to be finalized by the Network Steering Board in 2002. (See also, Background: page 2.)

Highlight 1. Draft TPA component list

- ❑ Definition of the trading partners
- ❑ Description of the purpose of the exchange
- ❑ Background information about the exchange – why the exchange is necessary
- ❑ Description of the benefits of the exchange
- ❑ Definition of the partners' roles and responsibilities related to data stewardship and Environmental Program (interest) responsibilities
- ❑ Documentation of relevant authorities and policies
- ❑ Agreement details: Data access (what data, exchange schedule, data access method), Metadata, Standards, Data (available how and reconciliation process), Technology, and Security
- ❑ Financial arrangements
- ❑ Period of agreement and termination
- ❑ Data ownership and rights of the partners
- ❑ Points of contact: Primary contacts and Support contacts
- ❑ Approvals/Signatures

Highlight 1 provides a proposed list of components likely to be included in a TPA and specified in the pending TPA guidance. This list will be expanded as more formal TPA guidelines are developed. Although they will cover many components, TPAs are likely to be only a few pages long.

Existing TPAs

To date, there are only two signed Network TPAs: One between the State of Nebraska, Department of Environmental Quality and EPA Region 7; and one between the State of Mississippi, Department of Environmental Quality and EPA Region 4. Both TPAs are for facility data exchanges between the states and EPA's Facility Registry System.

9. How much does a node cost to develop, manage, and maintain?

Node development

Node development will be similar, but narrower in scope, to agencies' current efforts to put information on the Web. It will be a minor cost for managing information. Once established, incremental costs to establishing new information flows will be reduced.

The costs associated with node development will become clearer as more Network participants build nodes. However, there are so many ways to develop nodes that there will not be a simple formula for how much a node costs. The limited experience of the state environmental agencies in the Alpha and Beta Phases of the Node Pilot Project provides an estimate of $5-$30K and approximately 80 hours of staff time for initial node development. (These 80 hours also included time devoted to learning and understanding XML technology.) These costs covered the purchase of separate node hardware, the software and outside technical support that was used to establish the node, and the staff time needed to develop internal capacity to manage the node. It also does not include the cost of designing, building, maintenance of and entering data into information systems.

The results of the Network Node Pilot Project – Beta Phase will help to establish where cost savings can be realized from developing "technology templates" for particular flows and common technologies. While these draft "templates" will be updated with further efforts in 2002, they will help to refine node cost estimates, which will be included in the Beta Phase project reports (expected in February, 2002.)

It is important to remember that a node is defined by what it does rather than what it is. Small node software packages, typically called middleware, entail costs like those described above. This is one approach to node development – and the approach that is likely to be the most common for the near future. However, several other approaches to development are available – such as using the built-in capabilities of newer enterprise software packages (e.g., Oracle 9i) in place of specialized node software/middleware. The investments for these large enterprise software packages, which serve agency-wide information management needs, may not require additional expenses for node development. While the enterprise information investment is high, the node software investment can be zero to relatively small.

Ongoing node management and maintenance

With experience, it is expected that establishing future node flows will require minimal commitments. The same is true for long-term node management and maintenance. Rather, doing so will require a fraction of the resources currently needed to establish additional information flows. As the Network continues to develop, and more nodes and information flows are used on a regular basis, ongoing node management is likely to become institutionalized. Establishing a new data flow will continue to take time, just as managing data exchanges and transfers currently take staff time. However, as previously explained, the advantages of using a node for official and unofficial flows include marked efficiencies over existing data flows. As such, any time spent on establishing and managing information flows through a node will be time well spent and, compared to other options, time saved.

10. Will financial support be provided for node development?

As of this writing, EPA expects to provide support for node development as part of the Core Capacity-Building ("Category B") Grants for which guidance will be provided to states late 2001 or early 2002. These grants will be intended to assist state agencies in addressing priority internal information technology investments while constructing initial linkages to the Network. Among other grant criteria, state agency grant proposals will include a three-year transition plan that addresses critical steps and milestones for achieving participation in the Network including the development of an operational node.

11. How will information security be handled?

Security is important and complicated. The Network seeks to have a common model for security that can be implemented by all Network participants. In 2002, the Network Steering Board will be developing specific guidelines for Network security standards. However, it is expected that Participants will use four levels of information security for node information flows (the Network Node Pilot Project – Beta Phase will be testing security Levels 1-3):

Level 1: Public information that requires no authentication or certification of integrity will be available through the Internet on a public, non-secure website. Level 1 security was tested in Alpha Phase.

Level 2: Information that requires some additional level of authentication (i.e., that it is the state environmental agency that is submitting the data) and a higher level of integrity protection will be available through the Internet on a website that is secured using Secure Socket Layer (SSL).

Level 3: Information at this level requires bi-directional authentication and a higher level of confidentiality protected by SSL at the server level, and it requires users' digital certificates.

Level 4: Information protection that requires non-repudiation in addition to privacy, authentication, and data integrity will be protected by requiring a digital signature "affixed" to the data that can be validated at the time of acceptance of the information by the environmental agency or the external user.

Aside from the digital signature feature in Level 4, these security measures are standard features of most web-server software and exchanges.

12. How stable are node technologies? Will these technologies be replaced overnight by newer ones?

XML is a stable standard that was developed by the World Wide Web consortium in 1998. XML has become the standard for exchanges of information over the Internet (several examples of how XML is being used in the private and public sectors are provided in Appendix B). For many uses, such as information exchanges over the Internet, XML is increasingly replacing Hyper Text Markup Language (HTML). Plus, unlike HTML, XML has the potential to be used for an extremely wide variety of information exchanges (see Appendix B). However, HTML will continue to be used for purposes, such as website development, for which it is well suited.

Several technologies used in conjunction with XML are under development. These include Simple Object Access Protocol (SOAP), Extensible Stylesheet Language (XSL), and XML Schema (XSD). Several of these technologies, such as XML Schema (an XML language for describing and constraining the content of XML documents), and XSL (Extensible Stylesheet Language – a standard that allows users to convert XML data into different formats) have already become standard. These standard technologies are the

ones the Network is being built upon. Other technologies are still being rapidly developed and it will take time to determine whether they should be relied upon for Network purposes.

Many software development companies are quickly developing XML software and user-friendly XML tools. The demand for XML tools is growing quickly, and as such, software developers are rushing to create superior XML tools. Several companies, including Oracle, IBM, Microsoft, and Sun Microsystems, have developed sound XML tools. In fact, these companies' entire software "outlook" has been upgraded to incorporate XML.

The fundamental technologies needed for the development of nodes are secure today and will not be replaced overnight by newer technologies. Some specific XML-related technologies are still under development, but these are the minority. Building nodes at this time is not premature.

13. Will node functions be separate from existing state agency web servers?

Ensuring the proper security and node functionality as outlined in the TPAs is the only standard to which participants will be held accountable. As such, Network participants will be able to choose their approach to their nodes, such as the server on which to host their node. Some Network participants will choose to have their node functions operate on a designated node server, while others may have their nodes operating on servers that also perform other functions such as public access websites. Providing a separate designated web server for a node may help prevent potential security problems and potential incompatibilities with existing server functions. The Network Node Pilot Project – Beta Phase group will continue to explore this question and will provide a recommendation on this subject as part of the project's documentation.

14. Who will be accountable for node transaction problems?

The accountability for node and information transfer problems will be defined in the TPAs. Each partner will have assigned responsibilities for the operation of and exchange of information through a node.

15. What about "legacy" systems?

Network participants have the flexibility to use the information systems they have now or are in the process of designing – from legacy systems to "data warehouses" to fully-integrated enterprise systems. XML software tools are routinely used to integrate data from disparate legacy systems. Agencies may decide to connect their nodes to consolidated warehouses that are fed by legacy systems, to connect their nodes directly to those legacy systems, or to mix these two approaches. In most cases, nodes will have "read only" access to these systems (which considerably simplifies node design).

16. Is a state *expected* to build and use a node?

As outlined in the *Blueprint,* participation in the Network is optional. Therefore, there is no mandatory requirement that state agencies build nodes. However, as of this writing, creating a node or advancing participation in the Network in some other compatible fashion is expected required under the Category B capacity-building grant (see Question 10).

III. Network Node Pilot Project: A Brief Summary

In January 2001 the State/EPA Information management Workgroup (IMWG), together with the Co-Chairs of the Action Teams and the Environmental Data Standards Council (EDSC), decided to produce near-term visible results that demonstrate the advantages of the Network approach. The project to produce these results would be called the Network Node Pilot Project and would fit within the "Pilot Flows and Applications" category of high-priority projects decided upon by this IMWG/Co-Chair group.

The Network Node Pilot Project has the following goals:

- To "ground truth" the XML technology referenced in the *Blueprint*;
- To demonstrate that states have the capacity to develop nodes and to expand upon earlier pilots that demonstrated the use of static XML transfers and integration;
- To establish a connection and exchange data between state agencies' nodes and EPA's CDX;
- To establish full production nodes that will concur with formal TPAs;
- To validate the node transactions against registered DETs[2] that normalize state data to a single common neutral format; and
- To develop recommendations for node coordination to the Interim Network Steering Group (INSG).

Alpha Phase (spring – summer 2001)

A group of interested state environmental agencies that were already involved in the Small States Technical Assistance Initiative and previously facility-related efforts expressed interest in launching the Network Node Pilot Project – Alpha Phase. Representatives from the agencies in Nebraska, New Hampshire, Delaware, and Utah participated in this phase to develop "proof of concept" nodes that quickly demonstrated that XML could be used to build nodes as described in the *Blueprint*. Using a common piece of "middleware" to link their information systems, the prototype nodes could answer specific queries of their facility data with standardized XML responses. The Alpha Phase participants consider this phase a success in that it produced the first proof-of-concept node prototypes. Based on their experience, the Alpha Phase group identified specific technical and management issues that will be addressed in the Beta Phase of the project.

Beta Phase (fall 2001- winter 2001)

[2] DETs are the Data Type Definitions (DTDs) used in the Alpha Phase or the Schema being used for the Beta Phase.

INSG is supporting the Beta Phase of the project. A primary goal of the Beta Phase is to learn enough about node development and functionality to come closer to Version 1.0 Network node specifications that can be implemented starting in early 2002. Specific objectives of the Beta Phase are as follows:

1. **Identify recommendations for node design specifications.** To prepare for full Version 1.0 Network nodes in many state environmental agencies, it will be necessary to establish node design specifications. These recommended specifications will help to inform and standardize the development of future production nodes.
2. **Scope the development of a "node package."** It is possible that the most efficient way to establish nodes for Version 1.0 will be to develop a few alternative node "packages" comprised of all the necessary components (e.g., hardware, software, coding, and processes) needed to establish a node. However, node hardware and software choices will still be up to those building nodes, allowing for interoperability and enforcing the fact that node criteria and standards are performance based.
3. **Implement a fully-normalized (i.e., rigorously standardized) facility data exchange template.** The DET being used for the Beta Phase is a facility Schema developed by the Facility Data Action Team and EPA's Facility Registry System (FRS).
4. **Involve 2-4 more state agencies as appropriate and as resources allow.** In addition to the four states that participated in the Alpha Phase, three other states (Florida, New Mexico), and EPA (CDX) are participating in the Beta Phase.
5. **Test and resolve performance issues.** Scalability, speed of data transport, and related performance issues will be tested as part of the Beta Phase.
6. **Explore security issues and options.** Two levels of security (Level 1 using HTTP and Level 2 using HTTPS and SSL) are being tested as part of the Beta Phase.
7. **Link state environmental agency node data transfers through EPA's node (CDX).**
8. **Pilot and/or execute TPAs for additional nodes.**
9. **Automate validation of flows against a DET.** The facility data flows will be validated against the facility Schema (see objective #3) that will be located on the test Registry developed by EPA for the Network.
10. **Send production data from nodes to the FRS.**
11. **Test interoperability**. (This objective is considered a "stretch," to be reached if time and resources allowed.) Interoperability is the ability of software and hardware on multiple machines from multiple vendors to communicate. The Beta Phase is testing at least three vendors' tools, and as such will be able to provide some insights regarding interoperability.

Initial results from the Beta Phase will be provided by the end of 2001. Final project documentation is scheduled to be submitted to the INSG and IMWG by January 2002. For more information on the Pilot Project, please see the following pages and/or go to the *Network Node Pilot Project – Beta Phase* Conference Room on WISER (http://www.ecos.org/wiser/).

Appendix A
High-Level Node Operation Chart

Note: The chart below displays the basic operations of most Beta Phase nodes. In this example, the node is hosted on a separate server and uses middleware to perform most node functions. However, there are many approaches to node implementation, and these approaches include performing node functions on state agency production servers using recent enterprise software that can perform node functions without the assistance of specialized XML middleware. It is expected that those establishing nodes will choose to use some kind of software to perform node functions as hand-coding all node functions would require additional time, effort, and in many cases, costs.

Figure A1 provides a simplified overview of one (hypothetical) node's operation process and Figure A2 provides a somewhat more detailed overview of the same process.

As the Network Node Pilot Project – Beta Phase continues in the fall and winter of 2001-2002, additional details will be provided regarding node architecture, specifications, and recommended implementation standards and options. (See Part III for more information on this project.)

Figure A1: Simplified Overview of A Node's Operation Process

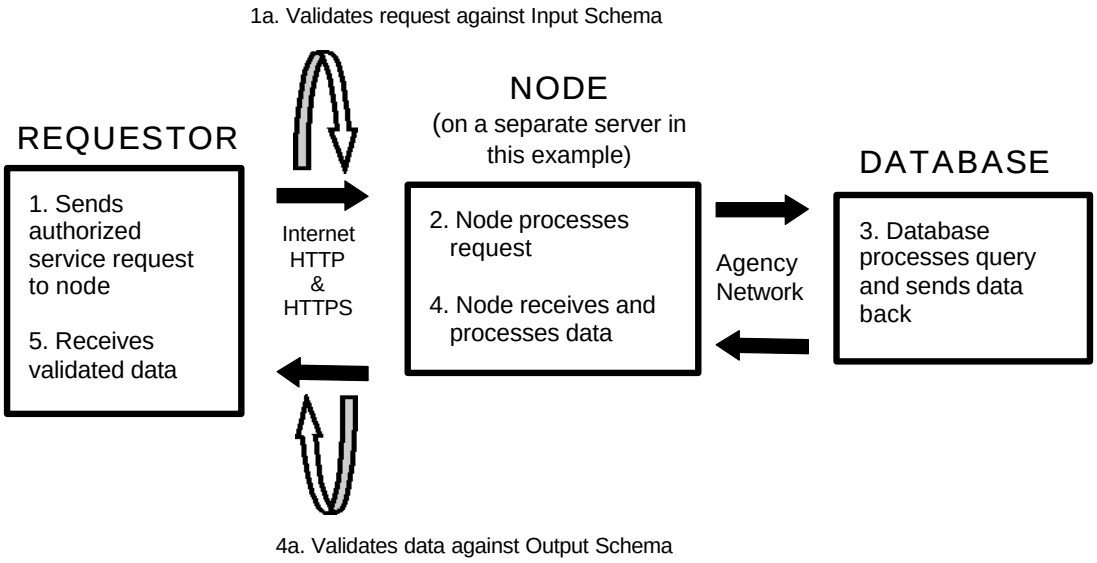
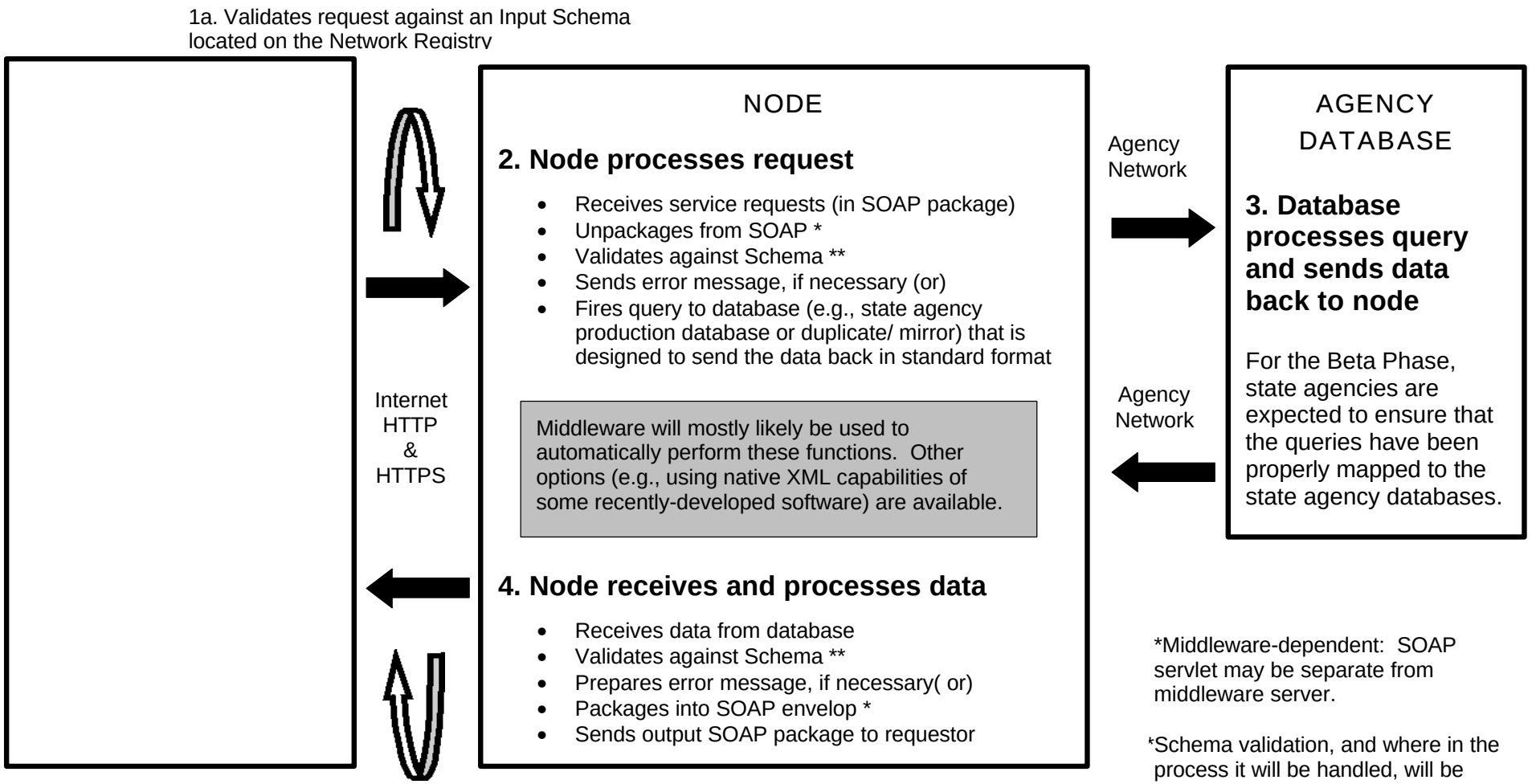


Figure A2: Overview of A Node's Operation Process



Appendix B

How Others are Using XML

United States Environmental Protection Agency (EPA) Current XML-Related Activities:

- EPA's Central Data Exchange (CDX) is preparing to accept National Pollutant Discharge Elimination System program (NPDES) data in XML through the Permit Compliance System (PCS), Interim Data Exchange Format. (http://www.epa.gov/cdx/idef.htm)
- EPA's Office of Solid Waste and Emergency Response (OSWER) is using XML for web database delivery functions related to the work of the Chemical Emergency Preparedness and Prevention Office Local Emergency Planning Committee Database. (http://www.epa.gov/ceppo/lepclist.htm)
- EPA participates in the Federal Government Content Network, which uses XML to integrate and manage government documents and information accessed by the public. (http://www.sdi.gov/server.htm)
- Air Emissions Inventory, from the office of Air and Radiation (OAR), is compiled annually with data reported by states to EPA. States were scheduled to begin submitting XML files to OAR in mid 2001.
- Unregulated Contaminant Monitoring Rule, administered in the Office of Water (OW), requires approximately 300 laboratories to submit test reports of drinking water to EPA. Several states submitted laboratory data in XML files.

United States Federal Aviation Administration (FAA): The FAA has many information systems that collectively provide services for air traffic separation, flight planning, weather information, and traffic flow management. With the increase of travel and the airline industry's shift to doing business over the Internet, the FAA realized the demand for fast access to detailed aviation data. The Adaptation Data Management Project addresses this demand by using XML to connect FAA information systems and provide consistent data. (http://www.gca.org/papers/xmleurope2000/pdf/s02-02.pdf)

United States Department of Energy – Los Alamos National Laboratory (LANL): LANL is a U.S. Department of Energy (DOE) laboratory located near Santa Fe, New Mexico. LANL's scientists work on national security issues ranging from biological research to solar system exploration. The Laboratory's computing systems, which include many simulation programs with specific input file formats and naming conventions, contribute to the Laboratory's reputation "as the greatest concentration of scientific computing power on the planet." To reduce the burden of managing large amounts of data and information, LANL uses XML to define and describe vocabularies from the different computer simulation programs. (http://www.gca.org/papers/xmleurope2000/pdf/s02-02.pdf)

United States Department of Energy – Office of Scientific and Technical Information (OSTI): OSTI uses XML for a majority of its web-based products, including website responses, search tools, and tracking website input. (http://xml.gov/scripts/efforts_detail.cfm?Effort=30)

United States Department of Defense (DOD): The DOD has developed a Registry to act as a clearinghouse for DOD-related industry and government coordination of XML, related technologies, and related data issues. (http://xml.gov/scripts/efforts_detail.cfm?EffortID=18)

United States Congress: The U.S. House of Representatives created XML.HOUSE.GOV, a public website that provides information related to House XML work and examples of XML use in government documents. In addition, Committees from the House and the Senate are working with the Library of Congress and the Government Printing Office to create Document Type Definitions (DTDs) for use in the creation of legislative documents using XML. (http://xml.coverpages.org/ni2001-08-13-a.html)

Health Level Seven (HL7) Standards Committee: The HL7 Standards Committee develops standards for the healthcare industry. Recently, the Committee established XML as the standard format for exchanging clinical documents and related healthcare information across many industry systems. (http://www.wedi.org/public/articles/details.cfm?id=324)

First Union National Bank: First Union uses XML documents for client application formats such as their commercial Cyberbanking application, credit scoring application, and home equity loan processing. (http://www.gca.org/papers/xmleurope2000/pdf/s02-02.pdf)

Appendix C
Glossary and List of Abbreviations

Blueprint – The *Blueprint for a National Environmental Information Exchange Network* (*Blueprint*) provides a conceptual design for the Network. The *Blueprint* document can be accessed at http://www.sso.org/ecos/eie/COMPLETE_BLUEPRINT_JUNE_01_FINAL.pdf.

CDX – EPA's Central Data Exchange (CDX). CDX is a centralized electronic report receiving system that will serve as EPA's enterprise-wide portal to the National Environment Information Exchange Network.

DET – Data Exchange Templates (DET) are empty but defined templates for data presentation and exchange. They identify what types of information are required for a particular document (i.e., name, address, etc.) as established in predefined standards or agreements. Specifically, DETs are typically either DTDs or Schemas.

Digital Certificate – A Digital Certificate, or Digital ID, is a means of verifying identity on the Internet. A third party company, known as a Certificate Authority, will research a group or individual and issue a Digital Certificate to them, vouching that they are who they say they are. This is a way to protect sensitive data by ensuring that others do not impersonate your site and accept or transmit data on your behalf.

DTD – A Document Type Definition (DTD) is an optional list of rules (markup declarations) to which a document or class of documents must conform, including information about what markup is valid in the document and the document's structure. A DTD can contain these rules internally, or link to them as an external subset. A DTD must be located before the first element of an XML document. It is generally referred to as DTD, but the abbreviation DTD can also mean document type declaration, depending on the context.

EDSC – Environmental Data Standards Council (EDSC) develops environmental data standards to promote the exchange of information among States, Native American Tribes, and EPA. The Council identifies those areas of information for which having standards will render the most value in achieving environmental results, prioritizes the areas, and pursues the development of data standards.

EPA – For the purpose of this document, EPA means all EPA offices (headquarters and regions)

Extensible – A language can be extended and adapted to meet many different needs

FRS – EPA's Facility Registry System (http://www.epa.gov/enviro/html/facility.html)

HTTP – HyperText Transfer Protocol (HTTP) is a protocol used to request and transmit files, especially webpages and webpage components, over the Internet or other computer network

HTTPS – A secure version of HTTP. Simply it is SSL underneath HTTP.

IMWG – State/EPA Information Management Workgroup (IMWG). Composed of senior leaders from EPA and state environmental agencies, the IMWG has initiated an approach to address joint information management in the form of a National Environmental Information Exchange Network (Network) (http://www.epa.gov/oei/imwg/)

INSG – Interim Network Steering Group (INSG). This group sunsets in January 2002.

Interoperability – The ability of software and hardware on multiple machines from multiple vendors to communicate

Middleware – a broad array of tools and data that help applications use networked resources and services

Network or The Network – The National Environment Information Exchange Network.

Node – A set of tools to exchange information on the Network. A node uses the Internet, a set of standard protocols, and appropriate security measures to issue and respond to authorized requests for specific information. Put another way, a Network node is a simple web service. A web service is software that exposes and executes the basic functionality of business applications through a standard Internet interface. A web service communicates with other web services via standards-based technologies that can be accessed by trading partners independent of hardware, operating system, or programming environment.

Node Pilot Project – The Alpha Phase was a pilot project conducted in the spring and summer of 2001. The project developed "proof of concept" Network nodes using XML technologies, demonstrating that these technologies could be used to build nodes as described in the *Blueprint*. Four states (Delaware, New Hampshire, Nebraska, and Utah) initiated and executed the pilot.

Node Pilot Project – The Beta phase, currently in progress, builds off the Alpha Phase and hopes to establish end-to-end flows with EPA, use a revised comprehensive DET for facility data, and develop draft specifications for node implementation.

NSB - Network Steering Board (NSB)

OASIS – Organization for the Advancement of Structured Information Standards. Their home site is http://www.oasis-open.org/. The DTD repository they sponsor is at http://www.XML.org.

Registry – (I don't know how to define this, do them mean something like the FRS or should this be defined conceptually

Servlet – A Java application that, different from applets, runs on the server and generates HTML-pages that are sent to the client. Servlets can run on browsers that are not Java-enabled.

Schema – A database-inspired method for specifying constraints on XML documents using an XML-based language. Schemas address deficiencies in DTDs, such as the inability to put constraints on the kinds of data that can occur in a particular field (for example, all numeric). Since schemas are founded on XML, they are hierarchical, so it is easier to create an unambiguous specification, and possible to determine the scope over which a comment is meant to apply.

SOAP – Simple Object Access Protocol (SOAP). SOAP is an XML/HTTP-based protocol for accessing services, objects and servers in a platform-independent manner.

SSL – Secure Sockets Layer (SSL) created by Netscape Communications, also known as secure server, provides for the encrypted transmission of data across the Internet. Users on both sides are able to authenticate data and ensure message integrity.

TPA – Trading Partner Agreement (TPA) defines the partners, information, stewardship, security, and other items essential for the exchange of information between two trading partners on the Network.

W3C – World Wide Web Consortium (W3C) is an industry consortium which promotes standards for the evolution of the Web and interoperability between WWW products by producing specifications and reference software. Although W3C is funded by industrial members, it is vendor-neutral, and its products are freely available.

Web Services – Web services are modular and reusable software components that are created by wrapping a business application inside a Web service interface. Web services communicate directly with other web services via standards-based technologies. These standards-based communications allow Web services to be accessed by customers, suppliers, and trading partners independent of hardware, operating system, or even programming environment.

XML – eXtensible Markup Langauage (XML) is a markup language defined by the W3C that provides a strict set of standards for document syntax while allowing developers, organizations, and communities to define their own vocabularies.

XSL– XSL is a language for expressing stylesheets. It consists of three parts: XSL Transformations (XSLT), a language for transforming XML documents; the XML Path Language (XPath), an expression language used by XSLT to access or refer to parts of an XML document; and XSL Formatting Objects, an XML vocabulary for specifying formatting semantics.

XSD – XML Schema Definition (XSD) language enables you to define the structure and data types for XML documents. An XSD schema defines the elements, attributes, and data types that conform to the W3C XML Schema Part 1: Structures specification for the XML Schema Definition language.

Appendix D
Additional Resources

General Network Background Materials

U.S Environmental Protection Agency, *Blueprint for a National Environmental Information Exchange Network*, Rev. 3 (June 2001).

U.S. Environmental Protection Agency, *National Environmental Information Exchange Network, Information Package* (June 2001).

XML and Related Technologies

Carlson, David. *Modeling XML Applications with UML, Practical e-Business Applications.* Addison-Wesley, 2001.

Cauldwell, Rajesh et al. *Professional XML Web Services*. Wrox, 2001.

Duckett, Jon et al. *Professional XML Schemas*. Wrox, 2001.

Hard, E.r., and W.S. Means. *XML in a Nutshell.* O'Reilly, 2001.

Van der Vlist, Eric. *XML.com: The Guide to W3C SML Schema.* O'Reilly & Associates, 2001.

Wyke, R. Allend, and Watt, Andrew. *XML Schema Essentials.* John Wiley & Sons, 2001.

Skonnard, Aaron, and Gudgin, Martin. *Essential XML Quick Reference: A Programmer's Reference to XML, Xpath, XSLT, SML Schema, SOAP, and More.* Addison-Wesley, 2001.

For more information about XML visit the following website at: http://www.xml.gov/

Additional Resources Referenced in this Document

W3C: http://www.w3.org/

OASIS: http://www.oasis-open.org/

ebXML: http://www.ebxml.org

The following references (URLs) were used to compile the glossary (Appendix B)

http://www.xmlmag.com/upload/free/features/xml/1999/01win99/glwin99/glwin99.asp

http://www.cyfi.com/products/dcfag.htm

http://www.dictionary.com/cgi-bin/dict.pl?term=interoperability&r=67

http://www.e-speak.hp.com/fag/fags_general.shtm#1

http://java.sun.com/products/isp/tomcat/

http://middleware.internet2.edu/overview/middleware-fag.shtml

http://www.softwareag.com/xml/about/glossary.htm

http://msdn.microsoft.com/library/

http://www.dictionary.com

http://www.netscape.com/info/